MENTAL HEALTH PREDICTION USING MACHINE LEARNING

Aarnna Salwan | Avani Todi | Krishna Maheswari | Chetak Kumar Mahato | Dr. I MalaSerene | SCORE

Introduction

Depression is a prevalent mental health disorder affecting millions worldwide. Early detection and classification of depression can help in timely intervention and treatment. With the rise of machine learning, it is now possible to predict and classify depression levels based on various factors such as stress, social support, financial stability, sleep, and physical activity. This study employs Support Vector Machine (SVM) and Random Forest Classifier to classify depression levels using real-world data. The dataset includes multiple psychological and lifestyle factors that influence mental health. The goal is to evaluate and compare the performance of these machine learning models in classifying individuals as having low or high depression levels.

A threshold score of 15 is set to classify depression into high and low categories, simplifying the prediction task. Correlation analysis is performed to examine feature relationships with depression scores, helping in selecting impactful variables. Machine learning models, including Support Vector Machine (SVM) with an RBF kernel and a Random Forest classifier, are trained to classify depression levels. These models are evaluated based on accuracy, precision, recall, and F1-score to measure their effectiveness. Additionally, confusion matrices are used to analyze classification errors, providing insights into the strengths and weaknesses of each model. Additionally, the models' performances are compared with and without feature selection to assess the impact of feature selection on prediction accuracy.

SCOPE of the project

This project aims to develop a classification system that can predict depression levels using machine learning techniques. The study focuses on extracting meaningful patterns from psychological and lifestyle data to improve the prediction accuracy. The models are trained using a dataset containing demographic details, mental health scores, and lifestyle parameters. The implementation of SVM and Random Forest allows us to analyze how different algorithms perform in terms of accuracy, precision, and recall. The results can be used to develop mental health assessment tools that can assist psychologists and medical professionals in early diagnosis and intervention.

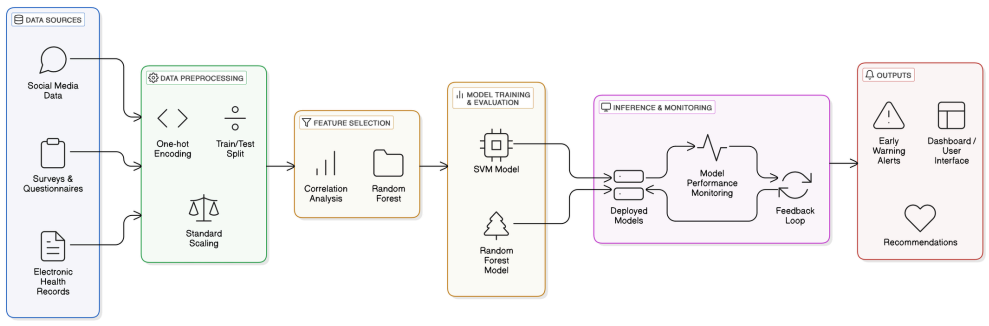
Methodology

The dataset used in this study consists of various psychological and lifestyle factors influencing mental health. To ensure data quality, missing values are checked and handled appropriately. Categorical variables such as gender, education level, employment status, and meditation habits are encoded using one-hot encoding to convert them into numerical values. Feature engineering is performed by introducing interaction features such as stress-financial interaction, support-loneliness interaction, and sleep-activity interaction, which help capture complex relationships between variables.

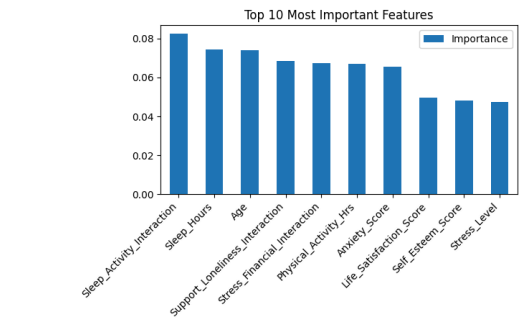
The dataset is then divided into training (80%) and testing (20%) sets using stratified sampling to maintain class balance. Standard scaling is applied to normalize numerical features, ensuring they have zero mean and unit variance, which enhances model performance.

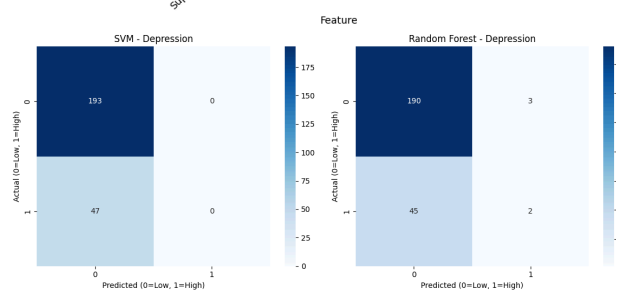
For feature selection, a Random Forest Classifier is used to determine feature importance, selecting the top ten most relevant features for depression prediction. Pearson correlation analysis is also conducted to understand the relationships between features and depression scores.

Architecture Diagram:



Results





The classification models show varying accuracy levels in predicting depression. The SVM model achieves an accuracy of 80.42%, while the Random Forest model achieves 80.00%. The confusion matrices indicate how well the models classify depression levels. The comparison of models with and without feature selection is also analyzed using bar graphs. The final results demonstrate the effectiveness of machine learning in mental health assessment.

Conclusion

Machine learning plays a crucial role in mental health analysis by providing efficient classification of depression levels. The study highlights that both SVM and Random Forest classifiers offer promising results, but feature selection significantly impacts the performance of these models. Further improvements can be made by incorporating deep learning techniques and larger datasets to improve predictive accuracy. This research can serve as a foundation for building automated mental health assessment tools for early diagnosis and treatment.

Contact Details

GitHub Repository: [Mental Health Prediction](https://github.com/Chetakk/Mental-Health-Prediction)
aarna.salwan2023@vitstudent.ac.in | avani.todi2023@vitstudent.ac.in |
krishna.maheswari2023@vitstudent.ac.in | chetakkumar.mahato2023@vitstudent.ac.in

Acknowledgments / References

Citations:
[1] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266516
[2] https://clinical-practice-and-epidemiology-in-mental-health.com/
[3] https://www.kaggle.com/datasets/ak0212/anxiety-and-depression-mental-health-factors
Thanks to VIT